Risk-limiting Audits for Nonplurality Elections

Anand Sarwate
UC San Diego

Stephen Checkoway
UC San Diego

Hovav Shacham
UC San Diego

Abstract

Many organizations have turned to alternative voting systems such as instant-runoff voting for determining the outcome of single-winner elections. It is our position in this paper that the increasing deployment of such alternative systems necessitate the study and development of risk-limiting audits for these systems. We initiate this study.

We examine several commonly used single-winner voting systems and provide risk-limiting auditing procedures for them. In many cases the methods from auditing plurality contests can be applied with minor changes and little loss in efficiency. For instant-runoff voting (IRV), the situation is markedly different. We describe an algorithm for auditing the candidate elimination order using plurality methods which is risk-limiting. Standard risk-limiting methods can be employed if the margin of the election can be efficiently calculated or bounded. We provide efficiently computable upper and lower bounds on the margin and, when known, compare them to the exact margins. Both auditing algorithms are potentially far less efficient than the methods to audit other types of voting systems.

1 Introduction

A post-election audit is a procedure that compares electronic vote tallies with paper ballots in an attempt to determine if the outcome is correct. For example in a single-winner elections, the audit is to decide if the winner is correct, not the tallies of the votes. A risk-limiting audit is one for which there is a known probability — the risk level — of certifying the reported outcome when it is incorrect. In this paper we describe how to perform risk-limiting audits for elections that use voting systems other than plurality or first-past-the-post.

California law requires a 1% manual tally of each election that compares the paper ballots to the machine records. Municipal elections in San Francisco use instant-runoff voting (IRV). For these IRV elections, the following manual tally procedure is employed [30]. First, in each randomly chosen precinct, the paper ballots are examined to determine the number of first-choice, second-choice, and third-choice votes each candidate received;[1] these totals are compared against the corresponding totals claimed in the original machine count. Second, an IRV elimination election is run with only the ballots from the tallied precinct, and the winner of this mini election is noted.

There is no reason to believe that the San Francisco tally of IRV elections is a risk-limiting audit for any particular risk level. Indeed, the San Francisco Voting Systems Task Force gives an example election in which two sets of ballots that are identical under the tally procedure produce two different election outcomes [30, Appendix A].[2] In this example election, running San Francisco's manual tally and finding no discrepancies does not increase our confidence that the reported and actual winner are the same! By contrast, San Francisco's 1% manual tally of a plurality election *does* provide a risk-measuring audit, though the risk level depends on the election margin.

Plurality voting has many well-documented drawbacks which has motivated the adoption of alternative voting systems. Proponents of these systems claim that voters can more clearly express their preferences and that the election winners better reflect the aggregate opinion of the electorate. Yet in the absence of a risk-limiting audit, it is difficult to argue that voters' preferences have been correctly aggregated — to convince the loser that (s)he lost.

It is our position in this paper that the increasing deployment of alternative voting systems such as IRV requires the development of risk-limiting audits for these alternative systems. We initiate the study of risk-limiting audits for a variety of alternative voting systems, by adapting techniques previously used for auditing plurality elections. An important ingredient in developing audits for these systems is to calculate the margin of victory and how errors in ballots affect the margin. In some cases calculating the margin is intuitive, but it is harder to understand the margin for IRV elections. There is likely room for substantial improvement over our audit techniques, and we hope that future work will provide these improvements.

In this paper, we are concerned only with ballot-based, risk-limiting audits. An audit is ballot-based if each ballot cast in the election can be independently selected and

[1] San Francisco allows voters to rank no more than three of the candidates for each race.

[2] In a presentation at the EVN 2011 conference, Emily Shen gave another such example.

compared to the corresponding paper ballot. A measure of an auditing algorithm's efficiency is the number of ballots that must be examined in the case that the reported outcome is correct. A simple risk-limiting audit compares every electronic record to its paper ballot. This is clearly inefficient since it requires examining each ballot.

Background: single-winner voting systems. An election is single-winner if exactly one candidate is declared the winner in each contest. In practice, not all elections are single-winner; for example, the three candidates with highest vote totals in a contest for city council might all capture seats on the council. We focus on single-winner elections in this paper to simplify the analysis, but our techniques might extend to multiple-winner elections.

The most common single-winner voting system in use is plurality voting — also called first-past-the-post or winner-takes-all. Well-known deficiencies in plurality voting have led to the development and use of other single-winner systems such as approval voting, Condorcet methods and instant-runoff voting (IRV). Arrow's impossibility theorem [1] and the Gibbard-Satterthwaite theorem [14, 31] tell us that no non-dictatorial election method based on preference ranking will simultaneously satisfy several desirable fairness criteria. The field of social choice theory is full of "paradoxes" of this nature that highlight difficulty in aggregating individual preferences into a group preference; more popular accounts can be found in recent books [25, 28, 38].

In the academic literature on post-election auditing, plurality voting has received the most attention, with very little work on simple variants such as range voting, and until very recently, almost no attention on elections using ranked-choices (Condorcet, IRV). In this paper we show that in many cases we can adapt or modify approaches from risk-limiting audits of plurality voting to other single-winner voting systems. However, in the case of IRV, the auditing problem is more difficult and our approaches are not as efficient as for other methods.

The apparent difficulty of auditing IRV is ironic. IRV is a popular alternative to plurality voting because it purports to avoid the spoiler effect and encourages many candidates to run. This can have policy benefits, as political parties can receive funding and recognition if they capture the first preferences of a certain portion of the electorate [27]. These benefits could be nullified if it is difficult to ascertain that an IRV election was correctly decided.

In this paper, we focus our attention on voting systems that enjoy some level of popularity. Other, less used, voting systems are considered when the problem of auditing such systems reduces to that of auditing more popular systems or when audit procedures can be obtained through minor modifications to other techniques.

Our contributions. In this paper, we propose post-election risk-limiting auditing schemes for several popular single-winner voting systems and variants.

We classify the voting systems we study into two groups: scored systems and ranked systems. We show that the first group, which includes approval voting, range voting, and Borda counts, can be audited using the analysis and algorithms from plurality election auditing. In some cases, we are able to reduce auditing systems in the second group to multiple-contest plurality elections. We can then apply the method of Stark [37] to audit the constituent plurality contests simultaneously with a given total risk of making a mistake. In particular, we give the first risk-limiting auditing algorithm for Condorcet methods (when there is a Condorcet winner) and for IRV.

2 Non-plurality voting and related work

We divide election methods for single-winner contests into two classes. These classes are distinguished by how the system *counts* a user's preference rather than how the user *expresses* that preference. The first class, *scored systems*, contains systems in which users' preferences are converted into a sequence of numerical scores for each candidate. The second class, *ranked systems*, contains systems in which comparisons are made based on the rank-ordering of candidates by the voters. Examples of the former include plurality voting, approval and range voting, and the Borda count. Examples of the latter include various Condorcet methods and instant-runoff voting.

Scored systems are perhaps the most familiar in modern day-to-day life. One might choose to "like" a piece of information posted on a social networking site, implicitly assigning a score of 1; and one often encounters restaurant reviews that assign ratings on a four-star scale. When electing candidates, as with deciding which piece of information is most popular or deciding which restaurant is best, the procedure is obvious: add up all of the scores and pick the alternative with the highest score.

Ranked systems are less common and the procedure for picking a winner is less straightforward. The social choice literature is replete with methods of aggregating preference rankings. Arrow's Theorem shows that no aggregation method can simultaneously satisfy a set of desirable criteria, which means there is no "perfect" method. The reasons for choosing one method over another involve evaluating tradeoffs between different notions of fairness or desiderata exogenous to the mathematics of the method.

Approval and range voting. In approval voting, voters simply mark "yes" or "no" for each candidate to indicate whether or not they approve of them. Range voting allows voters to assign a numerical score (e.g., 1 to 10) to their approval to indicate the degree to which they approve of each candidate. In both cases, the candidate with the highest number of points or "yes" votes wins. Since voters can approve of more than one candidate, advocates of approval and range voting claim[3] it avoids the "spoiler effects" rampant in plurality voting, wherein a third candidate siphons votes from one of the top two candidates.

Borda counts. Many of the classical results in voting date to the time of the French Revolution. In Borda count elections, voters submit their preferences in terms of a ranking of the candidates. Jean-Charles de Borda proposed a method of tallying ranked ballots by assigning points to each rank and giving these points to the candidates [6, 15]. In a 5 candidate election, a voter would give 5 points to their first-ranked candidate, 4 to the second-ranked, 3 to the third-ranked, and so on. Borda counts are used for some political elections in Slovenia [11], as well as the pacific island nations of Nauru and Kiribati [26], but is most popular in the worlds of sports — e.g., the Heisman Trophy [17] — and academic professional societies.

Condorcet methods. Condorcet methods are a class of voting systems in which voters rank candidates. For each pair of candidates i and j, the electorate is said to prefer i to j if more ballots rank i above j than vice versa. A *Condorcet winner*, if one exists, is a candidate i who is preferred to each other candidate. A tabulation method satisfies the *Condorcet criterion* if it elects the Condorcet winner when one exists. Any method that satisfies the Condorcet criterion is a Condorcet method.

It is possible to write endlessly on Condorcet methods. Indeed, a search through the literature will turn up a wide assortment of election mechanisms which satisfy the Condorcet criterion in addition to a host of other criteria. For descriptions and discussion of the most common Condorcet methods prior to 1980, including Condorcet's original, see Fishburn [12] and Tideman [39]. In the latter work, Tideman also describes his ranked pairs method where each pair of candidates is compared on every ballot. If there is a Condorcet winner, she is elected. Otherwise, the winner is chosen by considering the magnitudes of the victories in the pairwise elections. More recently, Markus Schulze proposed what has become the most commonly used Condorcet method, including by the Swedish Pirate Party for primaries, the Wikimedia Foundation, the Debian project, and the Gentoo project [33].[4]

Several Condorcet methods have been designed to satisfy additional desirable criteria beyond the Condorcet criterion such as the Condorcet loser criterion, the Smith criterion (also called the generalized Condorcet criterion), and the independence of clones criterion. The social choice literature has extensively analyzed these and other properties of voting systems. See Fishburn [12], Woodall [40], Schulze [32, 33] and the references therein for details.

Instant-runoff voting. In instant-runoff voting (IRV) — sometimes called the alternative vote (AV), ranked choice voting (RCV), or, for the multiseat version, single transferable vote (STV) — voters also submit their preferences as a (possibly truncated) ranked list of the candidates. It is probably the most widely used non-plurality voting system for political elections. The Australian House of Representatives uses STV [9], as does the Republic of Ireland for all public elections including presidential elections and elections to Dáil Éireann — the lower house of parliament [11]. The California cities of Berkeley, Oakland, San Francisco, and San Leandro use IRV for some elections. In an IRV election, candidates are eliminated sequentially, beginning with the candidate receiving the fewest first-ranked votes. The ballots whose first-ranked candidate was eliminated are assigned to their second-ranked candidates. A more detailed description of IRV is given in Section 4.2.

The IRV elections in Australia use paper ballots where voters fill out an explicit ranking for all of the candidates. The ballots are twice counted by hand under the watch of "scrutineers" — the election observers [9]. In California, voters only choose their top three choices, no matter how many candidates are on the ballot.

California law mandates a 1% manual tally to validate the outcome of an election. For IRV elections, this involves selecting 1% of the precincts at random and for each precinct, counting the number of first, second, and third choices for each candidate, and then running a "mini-IRV" election using only the ballots for the precinct. This is not a risk-limiting procedure and that data produced by it (the tallies and within-precinct IRV winner) give no information about the election as a whole [30, Appendix A].

One particularly challenging problem we consider is computing the margin for an IRV election. Two recent papers consider this problem as well. One, by Cary [7], gives an algorithm that computes a lower bound for the margin. Another, by Margino et al. [21], gives an algorithm that for many real IRV elections is able to compute the margin exactly. Though these papers were independent of and simultaneous with our initial work, in this revised version we make use of Cary's lower bound, and (where it is available) of Magrino et al.'s exact margin. To these we add an algorithm for computing an upper bound for the margin.

[3] See, for example `http://rangevoting.org`.

[4] Wikipedia lists 60 organizations which use the Schulze Method in some form. `http://en.wikipedia.org/w/index.php?title=Schulze_method&oldid=434396935#Use_of_the_Schulze_method` Accessed 2011-06-15.

Risk-limiting audits. In this paper we are interested in the problem of auditing the outcome of an election. An audit is a procedure in which samples of the ballots cast in the election are drawn and compared to the electronic record of the ballot used in the tally. The purpose of the audit is to look for evidence that the reported winner is not actually the winner. If the procedure does not find enough such evidence, the reported winner is declared the true winner. An audit procedure is called risk-limiting at risk level α if the probability (over the sampling procedure) that it incorrectly confirms the reported winner is less than α. Our underlying assumption is that some of the ballots may be miscounted due to human or machine error.

There has been a great deal recent work on performing risk-limiting audits. See Checkoway et al. [8], Stark [37], and the references therein for a thorough discussion.

We define the *margin* of an election to be twice the number of erroneous ballots needed to change the winner of an election. In the case of plurality elections, this is the gap between the top two candidates when the that gap is even, and one more than that when the gap is odd[5] This margin can be computed for alternatives to plurality voting as well. For example, consider an approval vote election between two candidates. If all voters approve of the first candidate but only half approve of the second candidate, then the old margin is $n/2$. However, $n/2$ votes need to be changed to change the outcome so the new margin is n. For auditing purposes, we think this is the right definition since it correctly mirrors the notion of detecting an appropriate number of material errors in a risk-limiting audit [36].

An important consideration for ballot-based audits is that they be *efficient* in the sense that the number of recounted ballots required to certify the election should not be too large. Many jurisdictions require an automatic recount if the margin below a certain percentage of the votes cast. For larger margins, it is desirable to develop audits which require as few ballots as possible to certify (when the outcome is correct) while remaining risk-limiting (when the outcome is not).

Ballot errors and strategic voting. The auditing problem is different than the problem of *strategic voting* studied in social choice theory and elsewhere. Strategic voting refers to the problem of voters casting ballots which do not reflect their true preferences in the hopes of exploiting the structure of the ballot tabulation system. It is well known that many voting systems are susceptible to strategic voting [13, 14, 31]. Determining if strategic voting will work in a given election is easy in the case of plurality voting, Borda count, and Condorcet voting [4], but Bartholdi and Orlin [3] have shown that it is NP-complete for a voter to find a preference order to ensure the election of a particular candidate under STV. However, if the number of candidates is small, the computational complexity may be tolerable. More recent work by Conitzer et al. has studied strategic voting of many systems from the perspective of computational complexity [10].

Strategic voting arises because voters have an incentive to cast ballots that do not reflect their true preferences. However, from the auditor's perspective, the voters' true preferences are irrelevant; a post-election audit is concerned with making sure that the voters' expressed preferences are counted correctly. A common question with both strategic voting and auditing is the following: Given the ballots cast in an election, how large a subset must an adversary control in order to force a particular outcome of the election? From the perspective of strategic voting, this subset is a coalition of strategic voters. From the perspective of auditing, the subset is the minimum number of errors required to change the outcome of the election.

3 Scored systems

Some methods proposed for auditing elections based on plurality voting can be easily extended to single-winner elections in which voters's preferences can be interpreted as scores given to each candidate. Efficient risk-limiting audits can be achieved using the methods of Stark [37] or Checkoway et al. [8]. We illustrate our ideas by slightly generalizing the method of Stark below.

Let k denote the number of candidates in the election and let $[k] = \{1, 2, \dots, k\}$ denote the set of candidates. Let n be the total number of ballots cast in the election. The true value of ballot i is $\mathbf{x}_i = \big(x_i(1), x_i(2), \dots, x_i(k)\big) \in [0,1]^k$, where $x_i(j)$ is the score that voter i gives to candidate j. Ballot i is recorded as $\mathbf{y}_i \in [0,1]^k$. The true and reported outcomes are

$$\mathbf{P} = \sum_{i=1}^{n} \mathbf{x}_i, \qquad \mathbf{Q} = \sum_{i=1}^{n} \mathbf{y}_i. \tag{1}$$

The reported winner and runner-up are

$$w_r = \operatorname{argmax}_{j \in [k]} \{Q(j)\}, \tag{2}$$

$$l_r = \operatorname{argmax}_{j \in [k]} \{Q(j) : j \neq w_r\}, \tag{3}$$

whereas the actual winner and runner-up are

[5] There is actually a slight difference concerning ties and even vs. odd margins which we will ignore in what follows. Additionally, when the threshold percentage of votes a candidate needs to be elected is different from 50%, for example a when a supermajority is required, shifting a vote from the reported loser to the reported winner can change the margin by less than 2 [16]. The factor of 2 in the definition of the margin is simply to agree with the standard definition of a margin for a plurality election and can safely be ignored in everything that follows if one wishes to work directly with number of changed ballots.

$$w_a = \text{argmax}_j\{P(j)\}, \quad (4)$$

$$l_a = \text{argmax}_j\{P(j) : j \neq w_a\}, \quad (5)$$

We will frequently write w for w_r since the auditor only knows w_r. The reported and actual margins are

$$m_r = Q(w_r) - Q(l_r), \qquad m_a = P(w_a) - P(l_a). \quad (6)$$

Note that these are measured in actual votes, not fractions or percentages of the number of ballots cast.

A *uniformly sampled, ballot-based audit* consists of drawing K ballots uniformly from the set of n ballots. Let $\mathbf{Z}_t = (\mathbf{X}_t, \mathbf{Y}_t)$ denote the t-th ballot in the sample (that is, $\mathbf{Z}_t$ is ballot $(\mathbf{x}_i, \mathbf{y}_i)$ for some i). The relative overstatement between candidates w and j with respect to the reported margin for the t-th ballot is

$$e_t(w, j) = \frac{\big(Y_t(w) - Y_t(j)\big) - \big(X_t(w) - X_t(j)\big)}{Q(w) - Q(j)}. \quad (7)$$

For elections where votes are in $\{0,1\}^k$, this is 0 when there is no error, positive (either $1/\big(Q(w) - Q(j)\big)$ or $2/\big(Q(w) - Q(j)\big)$) when there is an error that decreases the margin, and negative (either $-1/\big(Q(w) - Q(j)\big)$ or $-2/\big(Q(w) - Q(j)\big)$) when there is an error that increases the margin. The expected value of $e_t(w, j)$ is

$$\mathbb{E}\big[e_t(w, j)\big] = \frac{\frac{1}{n}\big(Q(w) - Q(j)\big) - \frac{1}{n}\big(P(w) - P(j)\big)}{Q(w) - Q(j)}. \quad (8)$$

For the t-th audited ballot, the worst case relative overstatement is

$$\hat{e}_t = \max_{j \neq w} \frac{\big(Y_t(w) - Y_t(j)\big) - \big(X_t(w) - X_t(j)\big)}{Q(w) - Q(j)}. \quad (9)$$

This is conservative, as $\hat{e}_t$ may be positive if the t-th sampled ballot shrinks the relative gap between the winner and *any* other candidate j. The expectation of $\hat{e}_t$ is not easy to calculate, but it is sufficient to upper bound it. The numerator is at most 2 and the denominator is at least m_r, so $\hat{e}_t \leq 2/m_r$. The test procedure consists of sampling ballots and computing the test statistic

$$T(K) = \prod_{t=1}^{K} \frac{1 - (m_r/n)/(2\gamma)}{1 - \hat{e}_t m_r/(2\gamma)}. \quad (10)$$

The election can be certified with risk α if $T(K) < \alpha$. The parameter $\gamma > 1$ effectively shrinks the margin m_r (or inflates the error) which helps make the test statistic more robust.

The approach described above by Stark [37] was proposed in the context of auditing plurality contests with possibly more than one winner. It is easy to transform approval, range voting, and Borda counts into this framework, as we can show in the remainder of this section.

3.1 Approval voting

In approval voting, each voter can decide to approve or disapprove of each candidate. Therefore the ballots are $\mathbf{x}_i \in \{0,1\}^k$. The auditing method was originally designed to work for the setting where voters could approve of up to c candidates and there were c winners, so this is a simple extension for approving of up to k candidates with 1 winner.

3.2 Range voting

In range voting users can assign a score to each candidate. These scores are typically integers, say from 0 to 10. The winner is the candidate who garners the maximum sum score from the voters. For a range voting system with scores in the range 0 to S we can normalize by S so that each ballot is represented by $\mathbf{x}_i \in \{0, 1/S, 2/S, \ldots, 1\}^k$. The auditing algorithm can then be run as before. Note that for range voting the upper bound of $2/m_r$ on $\hat{e}_t$ may be significantly more conservative than for approval voting, especially if many voters do not have polarized views about all of the candidates. This decreases the efficiency of the audit since it uses more ballots than necessary.

3.3 Borda count

Borda counts is thought of as a voting system where users rank candidates. This is true in that users submit their preferences in terms of a ranked list. However, the Borda count converts this ranked list into a numerical score for each candidate, and hence can be audited by the same mechanism as other scored systems. On a ballot for an election to be tabulated by a Borda count, voters rank candidates in order of preference. In an election with k candidates, the Borda count assigns $k - s + 1$ points to the s-th highest ranked candidate. Thus the top-ranked candidate for the voter gets k points, the second-ranked candidate gets $k - 1$ points, and so on. Voters need not rank all candidates; an unranked candidate gets 0 points. Again, by dividing the number of points by k, we can represent the i-th ballot as $\mathbf{x}_i \in \{0, 1/k, 2/k, \ldots, 1\}^k$.

4 Ranked systems

As discussed in Section 2, some alternative voting systems ask voters to explicitly rank candidates. We showed earlier that the Borda count is best thought of as a scored system, but Condorcet and IRV elections use the ranking order in a fundamentally different way.

Consider an election with k candidates and n ballots cast. For a set $A \subseteq [k]$, let $\Pi(A)$ denote the set of all ordered subsets of $[k]$. That is, $\Pi(A)$ contains all ranked lists of elements of A. In a ranked-choice election with k candidates, a ballot $\mathbf{x}_i$ for voter i is an element of $\Pi([k])$. The ballot $\mathbf{x}_i$ is recorded as a ballot $\mathbf{y}_i$. The election systems we discuss in this section all operate on the counts of the election. For an $S \in \Pi(A)$ define

$$N(S) = \sum_{i=1}^{n} \mathbf{1}(\mathbf{y}_i = S). \tag{11}$$

That is, $N(S)$ is the number of ballots recorded as preference ranking S.

Unlike in scored systems, we cannot create a common framework for tabulating ranked systems, but the two methods we discuss in this section, Condorcet and IRV, perform simple arithmetic operations and comparisons on the ballots in order to compute the outcome of the election.

4.1 Condorcet methods

To tabulate a Condorcet election, the counts are converted into pairwise preferences

$$C(i,j) = \sum_{S \in \Pi([k])} N(S) \cdot \mathbf{1}(i \text{ precedes } j \text{ in } S). \tag{12}$$

That is, $C(i,j)$ is the number of ballots in which i is ranked higher than j. If there exists a candidate $i \in [k]$ such that $C(i,j) > C(j,i)$ for all $j \neq i$, then candidate i is called the *Condorcet winner*. The *Condorcet graph* has vertices which are the candidates and the directed edge from i to j with weight $C(i,j)$.

If there is a Condorcet winner we can audit each edge connecting the winner to the other candidates in the Condorcet graph by considering a plurality election between the two candidates. Certifying the Condorcet winner will then certify the election. Because we are concerned with the winner only, we need to simultaneously audit $k-1$ pairwise elections and do not need to consider the $\binom{k}{2} - k + 1$ other pairwise elections. Under this auditing procedure, the reported margin of an election with Condorcet winner w_r is

$$m_r = \min_{j \neq w_r} \{ C(w_r, j) - C(j, w_r) \}. \tag{13}$$

One way to audit these is to use Stark's method of auditing a collection of races simultaneously [37] using a *diluted margin* of $\mu = m_r/n$.

If there is no Condorcet winner, then there is a *majority rule cycle* and we need to consider the particular Condorcet completion method used. There is a veritable menagerie of Condorcet completion methods proposed in the literature. To illustrate how auditing applies, we restrict our discussion to a few examples for which auditing is simple to describe.

Two-method systems. A two-method system elects the Condorcet winner, if one exists. If there is no Condorcet winner, than a completely different method of tabulating the ballots is used. One possible completion method to use when there is no Condorcet winner, first described by Black [5], uses Borda count to decide the winner. Fishburn improves on this by restricting the Borda counts to the Smith set — the smallest set of candidates such that each beats all candidates outside the set [12, Function C_1].

Auditing a two-method system involves auditing each method. If the reported counts indicate a Condorcet winner we can audit at risk level α using the methods described above. If the reported counts indicate that there is no Condorcet winner we first audit ballots to assure that no Condorcet winner exists at risk level r_1 by simultaneously auditing the pairwise elections that form the majority rule cycle. We can then audit the completion method (e.g., Borda count) at risk level r_2. We pick r_1 and r_2 such that

$$1 - (1 - r_1)(1 - r_2) \leq \alpha. \tag{14}$$

One-method systems. A one-method system is a single procedure that elects the Condorcet winner when one exists, and selects a different candidate otherwise. For the same set of cast ballots, different one-method systems may elect different candidates. If there is a reported Condorcet winner, the election can be audited using either the general method above or by auditing the specific method used. If there is no reported Condorcet winner, then a secondary audit must be used for the specific method.

The Nanson method [22] and the related Baldwin method [2] work in rounds with one or more candidates eliminated each round, similar to instant-runoff voting, except that Borda counts determine who is eliminated. The auditing procedure is very similar to IRV (Section 4.2). The Schulze method — the most commonly used Condorcet method — is more complicated. Developing a risk limiting audit for the Schulze method is an open problem. However, most organizations which use the Schulze method do not use physical ballots or a voter-verifiable paper audit trail (VVPAT), so the auditing framework used here may not be appropriate.

4.2 Instant-runoff voting

In an IRV election, voters also express their preferences as a ordered subset of the candidates. The counting proceeds in rounds. In each round, the candidates with the fewest top-choice votes are eliminated. Eliminating a candidate effectively removes the candidate from all ballots in which she was ranked, causing later ranked candidates to move up one spot. A candidate who is not eliminated is called a *continuing* candidate. A ballot is considered *exhausted* when all of the candidates it ranks have been eliminated. The elimination stops when one candidate has a majority of top-choice votes on the nonexhausted ballots.

There are several methods for choosing the candidates to eliminate. The simplest is to eliminate the candidate with the fewest top-choice votes. This is the base IRV elimination rule. In San Francisco municipal, ranked choice voting (RCV) elections, multiple candidates can be eliminated in a single round.[6] We refer to this as the SF RCV elimination rule. In both cases, the sum of the top-choice votes for candidates chosen to be eliminated is less than the number of top-choice votes for every candidate who is not eliminated (except in the case of a tie). That is, if E is an elimination set, then

$$\sum_{i=1}^{n} \mathbf{1}\big(y_i(1) \in E\big) < \min_{c \notin E} \sum_{i=1}^{n} \mathbf{1}\big(y_i(1) = c\big) \qquad (15)$$

where $y_i(1)$ is the top, noneliminated choice on ballot i. Other elimination rules — for example, eliminating all candidates who do not receive a threshold fraction of the top-choice votes each round — exist but will not be discussed. These two counting method are standard, but are provided for completeness in Algorithm 4 of Appendix A.

Tabulating the outcome of an IRV election produces a list $\mathcal{E} = (E_1, E_2, \ldots, E_R)$ of sets of eliminated candidates in the order in which they were eliminated. The set E_r is the set of candidates eliminated in the r-th round. Under the base IRV rules, E_r is always a single candidate for $r < R$, whereas in the SF RCV rule E_r may contain many candidates. In either case, once one candidate has a majority, the final elimination set E_R may contain multiple candidates.

There are at least three approaches to designing risk-limiting audits for IRV elections.

Auditing the taint. Given the margin of the IRV election, we can try to evaluate how much each erroneous ballot contributes to the margin — the so-called taint [35] — much as Stark's method does for plurality audits. Suppose we sample a ballot $\mathbf{X}_t$ whose cast vote record is $\mathbf{Y}_t \neq \mathbf{X}_t$, for example, the ballot may have been misinterpreted by the voting system or an adversary may have caused the cast vote record to be changed or recorded incorrectly. If replacing $\mathbf{Y}_t$ by $\mathbf{X}_t$ does not change the (plurality) margin in any elimination decision of the IRV tabulation process, we can declare the ballot *nonmaterial*. If any margin changes, then we declare the ballot error *material*; however, it is not clear how much the ballot changes the margin without recomputing the exact margin with the corrected ballot — a potentially lengthy computation. Indeed, it may be that the effect of a collection of errors on the margin may not be the sum of the effects of each error. Determining how to efficiently compute the size of the change is an open problem, the solution to which may provide drastically more efficient auditing methods than the two approaches below.

Auditing the elimination order. A second approach is to audit the elimination order $\mathcal{E}$ to verify that the set of candidates eliminated in each round is correct. If any elimination selection is a result of a tie breaker — for example, with the base IRV elimination rule, if the two candidates with the fewest number of top-choice votes in a round have the same number of votes, then the candidate to be eliminated may be chosen by some other mechanism such as a coin flip — then a complete hand count is necessary.

Otherwise, each round of the algorithm leads to a plurality election to be audited. For each round r: (1) eliminate and distribute the votes for candidates eliminated in previous rounds, namely $E_1 \cup E_2 \cup \cdots \cup E_{r-1}$; (2) aggregate the candidates who are to be eliminated in round r, namely those in E_r, into a "super candidate"; and (3) audit a $(k'-1)$-winner plurality election with k' candidates consisting of the super candidate and the $k'-1$ continuing candidates. The audit in step (3) is to ensure that the super candidate lost. This procedure results in R plurality elections to audit.

The R plurality elections can be audited simultaneously using Stark's method [37]. Each ballot can cause 0, 1, or 2 errors for each of the R-plurality elections; however, due to the nature of the *diluted margin* in Stark's method, we take the maximum of the errors caused in any race as the error contributed by the ballot.

In theory, this auditing method solves the problem of performing a risk-limiting audit for IRV elections, but in practice it may require counting too many ballots using the base IRV rules. This is because candidates who are eliminated early often constitute a very small fraction of the total ballots. For example, in the 2010 Oakland Mayoral election, three candidates each received less than 1% of the votes. This led to a small margin of 83 votes in round 3 out of a total of 122,264 ballots cast in the election. Small pairwise margins for candi-

[6] S.F., CAL., CHARTER art. XIII, § 13.102(e) (Mar. 2002), "If the total number of votes of the two or more candidates credited with the lowest number of votes is less than the number of votes credited to the candidate with the next highest number of votes, those candidates with the lowest number of votes shall be eliminated simultaneously and their votes transferred to the next-ranked continuing candidate on each ballot in a single counting operation."

dates eliminated early-on in the counting requires large sample sizes to detect an error in the elimination order. If instead of the base elimination rule, the SF RCV rule is used, then 8 of 11 candidates are eliminated in the first round and the smallest margin used for the audit is 1,627. We will return to this example several times in Section 5.2.

Auditing by error detection. A third approach to building a risk-limiting audit is to attempt *error detection*. That is, the auditor can sample K of ballots and compare each paper ballot to its cast vote record (CVR). If the number of ballots with *any* error exceeds a specified threshold, then a manual count of the entire election is required. Suppose that the margin is m. The effect of auditing for material errors is to audit a fictitious plurality contest between two candidates whose margin is m. Each material error that is found reduces the margin by two. Therefore any method for auditing plurality contests may be adapted for the purposes of error detection. Such an audit can be performed using any of the standard methods [19, 23, 29]. If fewer erroneous CVRs are found than the threshold, the auditor certifies the winner of the election. We choose the threshold so that the sample-size is risk-limiting.

This last approach to risk-limiting audits requires computing the margin of an IRV election, which is a topic of recent interest [7, 21]. Once the margin or a lower bound on it is known, then we can set the threshold to guarantee a risk-limiting audit. Recent work by Magrino et al. calculates the IRV margin exactly for some elections [21]. However, this exact calculation can be computationally very expensive, even with clever heuristics.

5 The margin of an IRV election

In this section we investigate the problem of computing the margin for an IRV election. We first describe some real IRV elections and their features. We then show a lower bound on the margin based on picking elimination sets in each round in such a way as to maximize the difference in votes between the "super candidate" described in Section 4.2 and the continuing candidate with the fewest votes. In order to evaluate how good this lower bound is, we develop an algorithm that constructs a set of ballot errors that can alter the winner of an IRV election. This gives an upper bound which is often close to the lower bound in real elections. Our bounds are fast to compute, and when possible we compare our bounds to the exact margins reported by Magrino et al. [21].

Appendix B contains several toy examples showing unintuitive aspects of IRV elections, namely that just a few errors recording ballots for losing candidates can dramatically change the outcome of the election and that even when IRV elects the Condorcet winner, the IRV margin can be significantly smaller.

5.1 Margins for real elections

We purchased CVR data for six different elections that were conducted using ranked-choice ballots from OpenSTV[7]. The three 2002 Dáil Éireann elections — Dublin North, Dublin West, and Meath — are multiple winner STV elections which we include to stress our algorithms, not because they are representative of IRV elections. The others are IRV elections. CVR data for an additional 26 San Francisco Bay Area and Pierce County RCV elections were also collected from the corresponding municipalities' websites. A summary of the data is given in Table 1.

The three Dáil Éireann, two Burlington mayoral, and Takoma Park City Council special elections allowed voters to provide a complete ranking of all of the candidates on the ballot. The last three additionally allowed write-in candidates although in the case of the two Burlington mayoral elections, the write-in took the place of one of the candidates in the ranking. All of the California and Washington elections used ballots where voters pick their top three candidates, including write-ins.

One common feature of all of these elections is that they involve a relatively small number of ballots compared to state and national elections. As an extreme example, only 204 people voted in the election for the Takoma Park City Council. In such cases a full hand count is easy, and would certainly be risk limiting. Indeed, in Australia, IRV elections are counted by hand twice under the supervision of scrutineers [9]. After tabulating the results from these elections we were surprised to note that they share a more interesting common feature: The winner according to the IRV count was also a Condorcet winner for the election in every case except for the 2009 Burlington mayoral election. In Appendix B.2 we show that the IRV margin may be smaller than the Condorcet margin, even when both methods elect the same candidate.

Computing the IRV margins exactly can be a computationally difficult task for real elections that contain large numbers of candidates or allow voters to rank many candidate on the ballot [21]. Therefore, in the rest of this section we present lower and upper bounds on the margin and examine the bounds for the 32 elections. Where known (from [21]), we compare the bounds with the exact margin.

5.2 Lower bounds on the margin

One obvious lower bound on the margin in an IRV election is the difference in votes between the two candidates with the fewest top-choice votes in each round. Certainly if the number of ballots which are modified is not enough change any elimination decision, then the outcome must be correct. However, it is trivial to show that this lower bound is arbitrarily bad by considering an election in

[7] `http://www.openstv.org`

Table 1: Election data.

Election	Candidates	Ranks	Ballots	Condorcet winner
2002 Dáil Éireann, Dublin North†	12	12	43,942	✓
2002 Dáil Éireann, Dublin West†	9	9	29,988	✓
2002 Dáil Éireann, Meath†	14	14	64,081	✓
2006 Burlington Mayor	6‡	5	9,865	✓
2007 San Francisco Mayor	18	3	149,465	✓
2007 Takoma Park City Council special, Ward 5	4‡	4	204	✓
2008 Pierce County Assessor	7‡	3	312,771	✓
2008 Pierce County City Council, Dist. 2	4‡	3	43,661	✓
2008 Pierce County Executive	5‡	3	312,771	✓
2009 Burlington Mayor	6‡	5	8,984	
2009 Pierce County Auditor	4‡	3	159,987	✓
2010 Berkeley Auditor	2‡	3	45,986	✓
2010 Berkeley City Council, Dist. 1	5‡	3	6,426	✓
2010 Berkeley City Council, Dist. 4	5‡	3	5,708	✓
2010 Berkeley City Council, Dist. 7	4‡	3	4,862	✓
2010 Berkeley City Council, Dist. 8	4‡	3	5,333	✓
2010 Oakland Mayor	11‡	3	122,268	✓
2010 Oakland Auditor	3‡	3	122,268	✓
2010 Oakland City Council, Dist. 2	3‡	3	15,243	✓
2010 Oakland City Council, Dist. 4	8‡	3	23,884	✓
2010 Oakland City Council, Dist. 6	4‡	3	14,040	✓
2010 Oakland School Board Director, Dist. 2	2‡	3	15,243	✓
2010 Oakland School Board Director, Dist. 4	3‡	3	23,884	✓
2010 Oakland School Board Director, Dist. 6	2‡	3	14,040	✓
2010 San Francisco Board of Supervisors, Dist. 2	7‡	3	28,911	✓
2010 San Francisco Board of Supervisors, Dist. 6	15‡	3	25,057	✓
2010 San Francisco Board of Supervisors, Dist. 8	5‡	3	38,551	✓
2010 San Francisco Board of Supervisors, Dist. 10	22‡	3	20,550	✓
2010 San Leandro Mayor	7‡	3	23,494	✓
2010 San Leandro City Council, Dist. 1	4‡	3	23,494	✓
2010 San Leandro City Council, Dist. 3	2‡	3	23,494	✓
2010 San Leandro City Council, Dist. 5	3‡	3	23,494	✓

† These are multiseat STV elections that have been treated as IRV.
‡ Includes a single combined write-in candidate.
The Ranks column denotes how many candidates a voter was allowed to rank on the ballot.
There is a ✓ in the Condorcet winner column if the IRV procedure elects the Condorcet winner.

Table 2: Margin bounds from real elections using ranked-choice ballots.

Election	Lower bound margin	%	Exact margin	%	Upper bound margin	%	Condorcet margin	%
2002 Dáil Éireann, Dublin North	203	0.5	—	—	2,724	6.2	2,723	6.2
2002 Dáil Éireann, Dublin West	8	0.0	—	—	1,444	4.8	1,443	4.8
2002 Dáil Éireann, Meath	129	0.2	—	—	6,198	9.7	6,197	9.7
2006 Burlington Mayor	482	4.9	776	7.9	778	7.9	776	7.9
2007 San Francisco Mayor	68,060	45.5	—	—	101,676	68.0	101,674	68.0
2007 Takoma Park City Council special, Ward 5	36	17.6	—	—	38	18.6	36	17.6
2008 Pierce County Assessor	276	0.1	2,222	0.7	2,222	0.7	2,221	0.7
2008 Pierce County City Council, Dist. 2	4,014	9.2	4,014	9.2	4,016	9.2	4,014	9.2
2008 Pierce County Executive	4,054	1.3	4,054	1.3	4,056	1.3	4,054	1.3
2009 Burlington Mayor	253	2.8	—	—	254	2.8	588	6.5
2009 Pierce County Auditor	16,792	10.5	16,792	10.5	16,794	10.5	16,792	10.5
2010 Berkeley Auditor	30,711	66.8	30,712	66.8	30,712	66.8	30,711	66.8
2010 Berkeley City Council, Dist. 1	1,770	27.5	2,348	36.5	2,350	36.6	2,348	36.5
2010 Berkeley City Council, Dist. 4	777	13.6	1,034	18.1	1,034	18.1	1,033	18.1
2010 Berkeley City Council, Dist. 7	728	15.0	728	15.0	730	15.0	728	15.0
2010 Berkeley City Council, Dist. 8	1,011	19.0	1,756	32.9	1,758	33.0	1,756	32.9
2010 Oakland Mayor	2,025	1.7	2,026	1.7	2,026	1.7	2,025	1.7
2010 Oakland Auditor	33,045	27.0	34,162	27.9	34,164	27.9	34,162	27.9
2010 Oakland City Council, Dist. 2	4,349	28.5	4,350	28.5	4,350	28.5	4,349	28.5
2010 Oakland City Council, Dist. 4	131	0.5	4,658	19.5	4,658	19.5	4,657	19.5
2010 Oakland City Council, Dist. 6	3,653	26.0	5,206	37.1	5,206	37.1	5,205	37.1
2010 Oakland School Board Director, Dist. 2	9,660	63.4	9,660	63.4	9,662	63.4	9,660	63.4
2010 Oakland School Board Director, Dist. 4	7,089	29.7	7,240	30.3	7,240	30.3	7,239	30.3
2010 Oakland School Board Director, Dist. 6	9,651	68.7	9,652	68.7	9,652	68.7	9,651	68.7
2010 San Francisco Board of Supervisors, Dist. 2	258	0.9	—	—	260	0.9	258	0.9
2010 San Francisco Board of Supervisors, Dist. 6	1	0.0	—	—	1,338	5.3	1,337	5.3
2010 San Francisco Board of Supervisors, Dist. 8	3,552	9.2	—	—	3,554	9.2	3,552	9.2
2010 San Francisco Board of Supervisors, Dist. 10	2	0.0	—	—	306	1.5	286	1.4
2010 San Leandro Mayor	232	1.0	232	1.0	234	1.0	232	1.0
2010 San Leandro City Council, Dist. 1	6,262	26.7	6,262	26.7	6,264	26.7	6,262	26.7
2010 San Leandro City Council, Dist. 3	16,675	71.0	16,676	71.0	16,676	71.0	16,675	71.0
2010 San Leandro City Council, Dist. 5	1,484	6.3	1,484	6.3	1,486	6.3	1,484	6.3

The exact margins are taken from Magrino et al. [21].

Algorithm 1: Lower bound on IRV margin

inputs: candidates A, ballots B
outputs: lower bound on margin lb

```
Enqueue(∞, (A,B))
while true do
    p, (A,B) ← Dequeue()
    if |A| = 1 then
        lb ← p
        return
    ESets ← ValidEliminationSets(B)
    foreach E ∈ ESets do
        A′ ← A \ E
        p′ ← min{ p, min_{c∈A′} Σ_{i=1}^{n} 1(y_i(1) = c) − Σ_{i=1}^{n} 1(y_i(1) ∈ E) }
        B′ ← EliminateCandidates(B, E)
        Enqueue(p′, (A′,B′))
```

which there are two candidates who each receive exactly one vote.

The example of the 2010 Oakland Mayoral race mentioned in Section 4.2 shows that different choices of elimination sets can lead to different margins in the constructed plurality elections used for the audit. Instead of auditing the elimination order by considering the actual elimination sets dictated by the election rules as described in Section 4.2, one can choose different, valid elimination sets for each round in such a way as to maximize the margin of the constructed plurality elections. This also provides a lower bound on the IRV margin that can be used to audit the taint or by error detection. Considering the 2010 Oakland Mayoral election again, if the 7 lowest candidates are eliminated in the first round instead of the 8 chosen by the SF RCV rule, then the smallest margin used in the audit rises to 2025.

The idea to pick the best elimination sets to use comes directly from David Cary's IRV lower bound computation in simultaneous, independent work [7]. For concreteness, we describe the lower bound algorithm and its correctness below but refer the interested reader to Cary's work for a more complete treatment as well as an alternative implementation.

The obvious lower bound is twice the number of ballots necessary to change the order that candidates are eliminated. However, by definition of the elimination set, in any round, any valid choice of elimination set can be chosen and those candidates eliminated without changing the ultimate winner of the election (this is the basis of the SF RCV elimination rule). This is a relaxation on the order in which candidates must be eliminated to ensure the correct outcome. Consider the Oakland Mayoral election one final time. Since the 8 candidates with the fewest top-choice votes in the first round can be eliminated without changing the winner, it is immaterial in which order those candidates are eliminated so long as they are eliminated before any others.

Any sequence of valid elimination sets thus gives a lower bound on the margin as follows. If $\mathcal{E} = (E_1, E_2, \ldots, E_R)$ is a valid sequence of elimination sets, then

$$lb_{\mathcal{E}} = \min_{E \in \mathcal{E}} \min_{c \notin E} \sum_{i=1}^{n} \mathbf{1}\big(y_i(1) = c\big) - \sum_{i=1}^{n} \mathbf{1}\big(y_i(1) \in E\big) \quad (16)$$

is a lower bound (cf. (15)). If each $E_r \in \mathcal{E}$ consists of a single candidate, then $lb_{\mathcal{E}}$ is the obvious lower bound.

Since each valid $lb_{\mathcal{E}}$ is a lower bound, we can take the maximum over all valid $\mathcal{E}$ to arrive at the bound

$$lb = \max_{\text{valid } \mathcal{E}} lb_{\mathcal{E}}. \quad (17)$$

The bound lb can be efficiently computed by using a priority queue. The queue initially contains the set of ballots with an infinite priority. The main loop removes the set of ballots with the highest priority p. If all but one candidates have been eliminated, then the priority is returned as the lower bound. Otherwise, all valid elimination sets are computed. For each valid elimination set, a copy of the ballots is constructed, the candidates in the set are eliminated, and the new ballots are placed into the queue with priority p' where p' is the minimum of p and the difference in votes between the sum of top-choice votes for candidates in the elimination set and the top-choice votes for the continuing candidate with the fewest votes. This procedure is given in Algorithm 1. Since we are using a priority queue, once we reach a set of ballots for which all candidates but one have been eliminated, every other sequence of elimination sets leads to a lower bound that is no better.

Note that the sequence of elimination sets used to construct the lower bound in Algorithm 1 is the optimal set to use when performing an audit of an IRV election by considering plurality elections for each round as described in Section 4.2. It is trivial to modify Algorithm 1 to return the sequence of eliminations used.

The weakness in the lower bound is that it considers the slimmest margin in any elimination decision. However, the margin between two candidates in a given round can be quite low but the candidates together have too many votes to be grouped into an elimination set. For example, in the 2010 Oakland City Council, District 4 election, in round 5 — using the base IRV elimination rule — Melanie Helby had 3,017 top-choice votes and Daniel Afford had 2,886. Neither can be eliminated in an earlier round using a larger elimination set and they cannot be eliminated together leading to a lower bound of 133 which is about 0.5% of the total number of ballots cast in the election. By contrast, the exact margin is 5,658 or about 19.5% of the ballots cast [21].

5.3 Algorithmic upper bound on the margin

In this section we develop an algorithm that takes a set of CVRs and constructs a set of ballot errors that changes the winner of the IRV election. This gives an upper bound on the margin of the election. This upper bound, which is efficiently computable, is useful to bound how far the lower bound described in the previous section is from the exact value if the exact value is not known. As we will show shortly, in many real elections the bound computed by this algorithm agrees with the exact margin.

Our method, Algorithm 2, is based on calculating an upper bound on the margin for each possible alternative winner of the election. For a given alternative winner j, we calculate a sufficient number of errors required to make j the winner of the election. Because our algorithm is "greedy" in a sense, the total number of errors we calculate may be quite a bit larger than the minimum number of material errors needed to change the outcome.

The algorithm proceeds as follows: For an alternative j, we tabulate the IRV election round by round until j appears in the elimination set E. Let k be the continuing candidate with the fewest top-choice votes; let s be the number of top-choice votes for candidates in E; and let m be the difference between the number of top-choice votes for k and s. The goal is to change enough votes from continuing candidates to j to remove j from the elimination set. There are two cases. If E is the singleton containing j — so that $s = m$ — then shifting $\lfloor m/2 \rfloor + 1$ votes from k to j will give j enough votes to not be eliminated in this round. Otherwise, there are multiple candidates in E and shifting $\lfloor (s-1)/2 \rfloor + 1$ votes from k to j is enough to remove j from the elimination set due to the strict inequality in (15). Finally, we tabulate the IRV election with the modified votes and repeat the margin modification process until a different candidate is elected. Note that it need not be j who is elected, since it is sufficient that *any* candidate other than the reported winner win.

The greedy part of the algorithm comes from how we choose the ballots for k that are switched to ballots for j. The selection heuristic is given in Algorithm 3. The method changes ballots of the form $(k, \ldots)$ to ballots with first-choice equal to j. We can write the elimination order $\mathcal{E} = (E_1, E_2, \ldots, E_R)$. The intuition is that for j to win, she must defeat the candidates in each of the sets $E_1 \setminus \{j\}, E_2, E_3, \ldots, E_R$ and finally w_r. So the heuristic is to preferentially change ballots closest to the elimination order of the election. This corresponds to lexicographically ordering the cast vote records as a function of $\mathcal{E}$ (see the definition of σ in Algorithm 3).

There are other heuristics possible for selecting which ballots to shift. For example, since j must eventually defeat w_r, it may be better to change preferentially ballots of the form $(k, w_r, \ldots)$, thereby greedily reducing the margin between j and w_r. Another set of heuristics can be derived by looking at the Condorcet graph of the ballots in round r and greedily ordering the ballots to be changed by the Condorcet margin. Since any heuristic generating a set of errors that alter the outcome of the election is a valid upper bound on the margin, we could take the minimum of margins generated by Algorithm 2 with each ballot ordering.

5.4 Margin calculations for real elections

Table 2 shows the results of margin calculations for the 32 elections in Table 1. We show four margin calculations: the lower bound of Section 5.2, exact margins from Magrino et al. [21] (when possible), the upper bound of 5.3, and the margin corresponding to treating the election as a Condorcet election.

In some elections, the lower bound produces a margin which is less than 0.5%, which is the threshold for a recount[8] in many jurisdictions. Because the three 2002 Dáil Éireann elections were for multiple-winner STV elections, the small values for the lower bound may not be representative. However, for two of the San Francisco Board of Supervisors elections, the lower bound produced a margin that is essentially zero whereas the upper bound is a significant fraction of the number of ballots cast.

For those elections where the exact margin was calculated by [21], our upper bound is either exactly the same or within two ballots of the exact margin. The difference in two ballots is because we take the margin to be twice

[8] Cf., ALA. CODE §17-16-20 (2010) or FLA. STAT. §102.141 (2010).

Algorithm 2: Greedy upper bound on IRV margin

inputs: candidates A, ballots $B = \{\mathbf{y}_i \in \Pi(A)\}_{i=1}^n$
outputs: upper bound on margin $\hat{m}$

```
Winner, ElimOrder, _ ← IRV (A, |A|, B)
foreach j ∈ A \ {Winner} do
    (Winner′, A′, B′, ElimOrder′) ← (Winner, A, B, ElimOrder)
    e_j = 0
    while Winner′ = Winner do
        l ← index of E in ElimOrder′ such that j ∈ E
        _, _, B′ ← IRV (A′, l − 1, B′) ⟨ eliminate candidates who would be eliminated before j ⟩
        foreach E ∈ ElimOrder′(1 : l − 1) do
            A′ ← A′ \ E ⟨ remove eliminated candidates ⟩
        k ← the candidate in A′ \ ElimOrder′(l) with the fewest top-choice votes
        s ← ∑_{c∈ElimOrder′(l)} (top-choice votes for c)
        m ← (top-choice votes for k) − s
        if |ElimOrder′(l)| > 1 then
            m ← m − 1 ⟨ modifying the margin by exactly m is enough to change the elimination set ⟩
        B′, e ← ModifyMargin (B′, m, j, k, ElimOrder′(l : end), Winner)
        e_j ← e_j + e
        Winner′, ElimOrder′, _ ← IRV (A′, |A′|, B′)
m̂ ← 2 min_j {e_j}
```

Algorithm 3: `ModifyMargin` – choosing errors to decrease the margin

inputs: ballots $B = \{\mathbf{y}_i \in \Pi(A)\}_{i=1}^n$, margin m, recipient j, victim k, ElimOrder, Winner
outputs: modified $\{\mathbf{y}_i\}$ such that the margin between j and k has decreased by more than m, number of ballots changed c

$c \leftarrow \lfloor m/2 \rfloor + 1$
$\sigma \leftarrow (\text{ElimOrder}(2:\text{end}) \parallel \text{Winner} \parallel \text{ElimOrder}(1) \setminus \{j\} \parallel j)$ ⟨ *σ is an ordered list* ⟩
$\sigma \leftarrow (k) \parallel \sigma \setminus \{k\}$ ⟨ *Move k to beginning of σ* ⟩
Sort $\{\mathbf{y}_i\}$ lexicographically according to σ, where longer matches appear before shorter matches; e.g., $(\sigma(1), \sigma(2), \sigma(3))$ precedes $(\sigma(1), \sigma(2))$
Change the first c of the sorted $\mathbf{y}_i$ into votes with j as the only choice

the number of ballot errors necessary for a different candidate to win, whereas [21] considers the number needed to tie. The tightness of our upper bound suggests that the heuristic that we used is an good one. However, for all of these elections the exact margin and upper bound are almost identical to the Condorcet margin. Thus in most real IRV elections, we suspect that the reported winner is likely to be the Condorcet winner and the margin is the Condorcet margin.

Many of the elections studied have margins which are quite large. This suggests that these elections should be efficiently auditable. However, these are not, in many cases, hotly contested elections. The 2010 Oakland School Board Director for District 6 was the only candidate on the ballot. It is unclear if such large margins would remain in larger, more contested elections such as state elections.

6 Conclusions

Alternatives to plurality voting like instant-runoff voting, if deployed, should be accompanied by risk-limiting audits. In this paper, we have initiated the study of risk-limiting audit procedures for these alternative single-winner election systems

We classify the alternative voting systems we study into two groups: scored and ranked systems. Scored voting systems can be audited using the analysis and algorithms from plurality election auditing. However, some ranked systems require a different approach. Condorcet elections where there is a Condorcet winner can be converted into a multiple-context plurality election. For instant-runoff voting, the question appears to be more complicated.

We propose three methods for auditing IRV elections. Auditing the taint requires an efficient method of calculating the margin (or a lower bound) and a way to compute how much an error in recording a ballot affects the margin — the ballot's taint — an open problem. Auditing the elimination order constructs a set of multi-winner plurality elections based on an optimal choice of elimination sets used to eliminate candidates in each round and audits each of those. Auditing by error detection is similar to auditing the taint except that all errors are assumed to be as bad as possible and modify the margin by 2.

Except for auditing by error detection — which is essentially the original ballot-based method for auditing plurality election — knowledge of the margin for an IRV election is not sufficient for auditing.

An important question is whether the effects seen in the relatively small elections for which we have data will be present in larger state or national elections. Obtaining data from such elections (or polling data) could be quite valuable. As with plurality elections, there is a dearth of information on the nature and distribution of real ballot errors. Such data could be used to optimize the statistical efficiency of an auditing procedure.

We do not believe that our proposed auditing procedures are the last word on risk-limiting audits of alternative election systems; we hope that future work will provide simpler and more efficient audits. As things stand now, however, some voting systems (e.g., Borda counts) appear to be substantially easier to audit than others (e.g., IRV). We believe that ease of auditing should be a criterion when voting systems are considered for adoption.

Acknowledgments

We thank Eric Rescorla and Joseph Lorenzo Hall for numerous helpful discussions and David Cary, Thomas R. Magrino, Ronald L. Rivest, Emily Shen, and David Wagner for sharing advance copies of their papers. This material is based on work supported in part by the National Science Foundation under Grant No. CNS-0831532, by the MURI program under AFOSR Grant No. FA9550-08-1-0352, and by the California Institute for Telecommunications and Information Technology (CALIT2) at UC San Diego.

Bibliography

[1] Kenneth J. Arrow. *Social Choice and Individual Values*. Yale University Press, second edition, 1951.

[2] Joseph M. Baldwin. The technique of the Nanson preferential majority system of election. *Proceedings of the Royal Society of Victoria*, n.s. 39:42–52, 1926.

[3] John J. Bartholdi III and James B. Orlin. Single transferable vote resists strategic voting. *Social Choice and Welfare*, 8(4):341–354, 1991.

[4] John J. Bartholdi III, Craig A. Tovey, and Michael A. Trick. The computational difficulty of manipulating an election. *Social Choice and Welfare*, 6(3):227–241, 1989.

[5] Duncan Black. *The Theory of Committees and Elections*. Cambridge University Press, 1958.

[6] Jean-Charles de Borda. Mémoire sur les élections au scrutin. *Mémoires de l'Académie Royale des Sciences*, pages 657–65, 1781.

[7] David Cary. Estimating the margin of victory for instant-runoff voting. In Shacham and Teague [34]. To appear.

[8] Stephen Checkoway, Anand Sarwate, and Hovav Shacham. Single-ballot risk-limiting audits using convex optimization. In Jones et al. [20].

[9] Australian Electoral Commission. Counting the votes – house of representatives. `http://www.aec.gov.au/Voting/counting/vid_hor.htm`.

[10] Vincent Conitzer, Tuomas Sandholm, and Jérôme Lang. When are elections with few candidates hard to manipulate? *Journal of the ACM*, 54(3), 2007.

[11] ElectionGuide. `http://www.electionguide.org/`.

[12] Peter C. Fishburn. Condorcet social choice functions. *SIAM Journal on Applied Mathematics*, 33 (3):469–489, November 1977.

[13] Peter Gardenfors. Manipulation of social choice functions. *Journal of Economic Theory*, 13(2):217–228, 1976.

[14] Allan Gibbard. Manipulation of voting schemes: A general result. *Econometrica*, 41(4):587–601, July 1973.

[15] Alfred de Grazia. Mathematical derivation of an election system. *Isis*, 44(1/2):42–51, June 1953.

[16] Joseph Lorenzo Hall. On the margin: The effects of introducing or swapping votes on election margins. `http://josephhall.org/eamath/margins09.pdf`, July 2009.

[17] Heisman Trophy Balloting. `http://www.heisman.com/history/balloting.php`.

[18] I. David Hill. An odd feature in a real STV election. *Voting Matters*, 18:9, June 2004.

[19] Kenneth C. Johnson. Election certification by statistical audit of voter-verified paper ballots. `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=640943`, 2004.

[20] Doug Jones, Jean-Jacques Quisquater, and Eric Rescorla, editors. *Proceedings of EVT/WOTE 2010*, August 2010. USENIX, ACCURATE, and IAVoSS.

[21] Thomas R. Magrino, Ronald L. Rivest, Emily Shen, and David Wagner. Computing the margin of victory in IRV elections. In Shacham and Teague [34]. To appear.

[22] Edward J. Nanson. Methods of election. *Transactions and Proceedings of the Royal Society of*

Victoria, 18(954):197–240, 1882.

[23] C. Andrew Neff. Election confidence: A comparison of methodologies and their relative effectiveness at achieving it (revision 6). Online: `www.verifiedvoting.org/downloads/20031217.neff.electionconfidence.pdf`, December 2003.

[24] Jeffrey O'Neill. Fast algorithms for counting ranked ballots. *Voting Matters*, 21:1–5, March 2006.

[25] William Poundstone. *Gaming the Vote: Why Elections Aren't Fair (and What We Can Do about It)*. Hill and Wang, 2008.

[26] Benjamin Reilly. Social choice in the south seas: Electoral innovation and the borda count in the pacific island countries. *International Political Science Review*, 23(4):355–72, October 2002.

[27] Sam Roberts. Minor parties see threat in ballot quirk. *The New York Times*, October 25, 2010.

[28] Donald D. Saari. *Decisions And Elections: Explaining The Unexpected*. Cambridge University Press, 2001.

[29] R. G. Saltman. Effective use of computing technology in vote-tallying. Technical Report Tech. Rep. NBSIR 75–687, National Bureau of Standards (Information Technology Division), Washington, D.C., USA, March 1975. `http://csrc.nist.gov/publications/nistpubs/NBS_SP_500-30.pdf`.

[30] San Francisco Voting Systems Task Force. Recommendations on voting systems for the city and county of San Francisco (draft). Online: `http://www.sfgov2.org/Modules/ShowDocument.aspx?documentid=155`, January 2011.

[31] Mark Allen Satterthwaite. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting and social welfare functions. *Journal of Economic Theory*, 10(2):187–217, April 1975.

[32] Markus Schulze. Condorect sub-cycle rule. The Election-Methods mailing list, October 1997. `http://lists.electorama.com/pipermail/election-methods-electorama.com/1997-October/001570.html`.

[33] Markus Schulze. A new monotonic, clone-independent, reversal symmetric, and condorcet-consistent single-winner election method. *Social Choice and Welfare*, 36(2):267–303, February 2011.

[34] Hovav Shacham and Vanessa Teague, editors. *Proceedings of EVT/WOTE 2011*, August 2011. USENIX, ACCURATE, and IAVoSS.

[35] Philip B. Stark. Conservative statistical post-election audits. *Ann. Appl. Stat.*, 2(2):550–581, March 2008.

[36] Philip B. Stark. Risk-limiting postelection audits: Conservative *P*-values from common probability inequalities. *IEEE Transactions on Information Forensics and Security*, 4(4):1005–1014, December 2009.

[37] Philip B. Stark. Super-simple simultaneous single-ballot risk-limiting audits. In Jones et al. [20].

[38] George Szpiro. *Numbers Rule : The Vexing Mathematics Of Democracy, From Plato To The Present*. Princeton University Press, 2010.

[39] T. Nicolaus Tideman. Independence of clones as a criterion for voting rules. *Social Choice and Welfare*, 4(3):185–206, September 1987.

[40] Douglas R. Woodall. Monotonicity of single-seat preferential election rules. *Discrete Applied Mathematics*, 77(1):81–98, June 1997.

A IRV tabulation algorithm

One way to perform IRV tabulation is described in Algorithm 4. It takes as input the set of candidates A, the maximum number of rounds of the algorithm to perform ρ, and the set of ballots $\{\mathbf{y}_i \in \Pi(A)\}_{i=1}^{n}$. It iteratively eliminates the candidates with the fewest top choice votes — $\mathbf{y}_i(1)$ is the top, noneliminated choice on ballot i, and then removes the candidates from every ballot on which they appear. As output, it produces the winner, if any after ρ rounds, the set of candidates eliminated in each round, and the modified set of ballots after candidates have been eliminated.

As discussed in Section 4.2, there several rules for choosing which candidates to eliminate in each round. By abstracting the choice of the elimination set, all varieties of IRV can be described at once. The function `EliminationSet(`B`)` in Algorithm 4 takes a set of ballots B and returns the set of candidates to be eliminated next. For example, using the base IRV elimination rule, `EliminationSet(`B`)` returns the candidate with the fewest top-choice votes. The SF RCV elimination rule returns largest set of candidates E such that the sum of the top-choice votes for all candidates in E is less than the number of top-choice votes for all of the candidates not in E, S.F., CAL., CHARTER art. XIII, § 13.102(e) (2002).

Rather than iterating over each ballot every time, one can pick smarter representations such as keeping track of how many ballots with each particular candidate ranking exist or using tree data structure in which paths from the root to a node correspond to candidate rankings [24]. Using a tree, eliminating a candidate involves recursively removing nodes corresponding to that candidate and merging their children.

B Examples for IRV margins

In this appendix we give some toy examples of IRV elections that illustrate two points. First, a small number of

Algorithm 4: `IRV` – tabulating IRV results

inputs: candidates A, rounds ρ, ballots $\{\mathbf{y}_i \in \Pi(A)\}_{i=1}^{n}$
outputs: Winner, ElimOrder, modified $\{\mathbf{y}_i\}$

Winner $\leftarrow 0$
ElimOrder $\leftarrow ()$
$r \leftarrow 0$
while $r < \rho$ **do**
  $r \leftarrow r+1$
  foreach $c \in A$ **do**
    $Q(c) \leftarrow \sum_{i=1}^{n} \mathbf{1}\big(\mathbf{y}_i(1) = c\big)$
  if $\max_c \{Q(c)\} > \frac{1}{2}\sum_c Q(c)$ **then**
    Winner $\leftarrow \mathrm{argmax}_c \{Q(c)\}$
    Append $A \setminus \{$Winner$\}$ to ElimOrder
    break
  else
    $E \leftarrow$ `EliminationSet`$(\{\mathbf{y}_i\})$
    Append E to ElimOrder
    $A \leftarrow A \setminus E$
    foreach $\mathbf{y}_i$ **do**
      $\mathbf{y}_i \leftarrow \mathbf{y}_i \setminus E$

errors can dramatically change the outcome of an IRV election. Secondly, the IRV margin can be smaller than the Condorcet margin, even when IRV elects the Condorcet winner.

B.1 IRV can be sensitive to small errors

IRV is sensitive to errors, in the following sense: Switching even a single vote from one losing candidate to another (or fabricating a vote for a losing candidate) may be enough to change the winner of an election.[9] We illustrate this via a simple example. Consider the six candidate, 1000 ballot election in Table 3. Zoë has the fewest votes of any candidate. She is eliminated in the first round and ultimately Velma wins with 496 votes. Ulric comes in second with 379 votes. Naïvely, one might say that Velma won with a margin greater than 10% (either 11.7% or about 13.4% depending on whether the denominator is 1000 or $379+496 = 876$).

If an adversary is able to arrange for a single Y X V ballot to be counted as a Z Y ballot, then we get the election in Table 4. Here, the small error cascades through the rest of the rounds and Ulric, who previously came in second, is the winner with 379 votes. The correct winner, Velma, does not even make it to the final round. Instead, Xavier, who was previously eliminated in the second round makes it all the way to the final round to lose with 295 votes. Again, naïvely, the margin appears to be quite large (either 8.4% or about 12.5%). This example shows that intuition about margin calculations in plurality elections may not be applicable to IRV elections.

B.2 Margins for Condorcet versus IRV

The margin for IRV may be smaller than the Condorcet margin. Consider an election between Xavier, Yolanda, and Zoë. Only 36 ballots were cast in this election, and the results are summarized in Table 5.

Under IRV, in the first round Xavier gets 11 votes, Yolanda 15 votes, and Zoë 10 votes, so Zoë is eliminated. However, the supporters of Zoë break unanimously for Xavier over Yolanda, so in the final round Xavier defeats Yolanda 21 votes to 15 and Xavier is the IRV winner. The simple lower bound for the margin of this election is one vote, the gap between Xavier and Zoë in the first round. Note that Xavier is also the Condorcet winner of this election — voters prefer Xavier to both Yolanda and Zoë by 21 to 15. The Condorcet margin is therefore six votes. Further, voters also prefer Yolanda to Zoë 21 to 15 so the minimum difference in preference between candidates is also six. However, the IRV margin really is two votes since one ballot shifted from Xavier to Zoë will cause Xavier to be eliminated in the first round and Yolanda to win.

[9] I. D. Hill describes a slightly different example of instability in a real Single Transferable Vote election — the multiseat analogue of instant-runoff voting. Hill points out that a change in a single ballot's 15th choice (out of 23) would result in a different winner. In this case, it was the difference between voting for one of the (eventual) winners and the closest runner up rather than between two losers [18].

Table 3: Unmodified six candidate, 1000 ballot IRV election.

Candidate	Round 1	Round 2	Round 3	Round 4	Final
Ulric	199: U	199: U	199: U	199: U 180: ~~W~~ U	199: U 180: ~~W~~ U
Velma	200: V	200: V	200: V 170: ~~X~~ V	200: V 170: ~~X~~ V	200: V 170: ~~X~~ V 126: ~~Y~~ ~~X~~ V
Wilard	180: W U	180: W U	180: W U	—	—
Xavier	170: X V	170: X V	—	—	—
Yolanda	126: Y X V	126: Y X V 125: ~~Z~~ Y	126: Y ~~X~~ V 125: ~~Z~~ Y	126: Y ~~X~~ V 125: ~~Z~~ Y	—
Zoë	125: Z Y	—	—	—	—

Table 4: IRV election in Table 3 with a single Y X V ballot changed to Z Y.

Candidate	Round 1	Round 2	Round 3	Round 4	Final
Ulric	199: U	199: U	199: U	199: U 180: ~~W~~ U	199: U 180: ~~W~~ U
Velma	200: V	200: V	200: V	200: V	—
Wilard	180: W U	180: W U	180: W U	—	—
Xavier	170: X V	170: X V 125: ~~Y~~ X V	170: X V 125: ~~Y~~ X V	170: X V 125: ~~Y~~ X V	170: X ~~V~~ 125: ~~Y~~ X ~~V~~
Yolanda	125: Y X V	—	—	—	—
Zoë	126: Z Y	126: Z ~~Y~~	—	—	—

Table 5: IRV election where the IRV margin is smaller than the Condorcet margin.

Candidate	Round 1	Final
Xavier	6: X Y Z 5: X Z Y	6: X Y ~~Z~~ 5: X ~~Z~~ Y 10: ~~Z~~ X Y
Yolanda	10: Y X Z 5: Y Z X	10: Y X ~~Z~~ 5: Y ~~Z~~ X
Zoë	10: Z X Y	—